Network Working Group
Request for Comments: 678
NIC: 31524

J. Postel
(SRI-ARC)
19 December 1974

Standard File Formats

Introduction

In an attempt to provide online documents to the network community we have had many problems with the physical format of the final documents. Much of this difficulty lies in the fact that we do not have control or even knowledge of all the processing steps or devices that act on the document file. A large part of the difficulty in the past has been due to some assumptions we made about the rest of the world being approximately like our own environment. We now see that the problems are due to differing assumptions and treatment of files to be printed as documents. We therefore propose to define certain standard formats for files and describe the expected final form for printed copies of such files.

These standard formats are not additional File Transfer Protocol data types/modes/structures, but rather usage descriptions between the originator and ultimate receiver of the file. It may be useful or even necessary at some hosts to construct programs that convert files between common local formats and the standard formats specified here.

The intent is that the author of a document may prepare his/her text and store it in an online file, then advertise that file by name and format (as specified here), such that interested individuals may copy and print the file with full understanding of the characteristics of the format controls and the logical page size.

Standardization Elements

The elements or aspects of a file to be standardized are the character or code set used, the format control procedures, the area of the page to be used for text, and the method to describe overstruck or underlined characters.

The area of the page to be used for text can be confusing to discuss, in an attempt to be clear we define a physical page and a logical page. Please note that the main emphasis of this note is to describe

the standard formats in terms of the logical page, and that it is up to each site to map the logical page onto the physical page of each of their devices.

Physical Page

The physical page is the medium that carries the text, the height and width of its area are measured in inches.

The typical physical page is a piece of paper eleven inches high and eight and one half inches wide.

Typical print density is 10 characters per inch horizontally and 6 characters per inch vertically. This results in the typical physical page having a maximum capacity of 66 lines and 85 characters per line. It is often the case that printing devices limit the area of the physical page by enforcing margins.

Logical Page

The logical page is the area that can contain text, the height of this area is measured in lines and the width is measured in characters.

A typical logical page is 60 lines high and 72 characters wide.

Code Set

The character encoding will be the network standard Network Virtual Terminal (NVT) code as used in Telnet and File Transfer protocols, that is ASCII in an eight bit byte with the high order bit zero.

Format Control

The format will be controlled by the ASCII format effectors:

Form Feed <FF>

Moves the printer to the top of the next logical page keeping the same horizontal position.

Carriage Return <CR>

Moves the printer to the left edge of the logical page remaining on current line.

Line Feed <LF>

Moves the printer to the next print line, keeping the same horizontal position.

Horizontal Tab <HT>

Moves the printer to the next horizontal tab stop.

The conventional stops for horizontal tabs are every eight characters, that is character positions 9, 17, 25, ... within the logical page.

Note that it is difficult to enforce these conventions and it is therefore recommended that horizontal tabs not be used in document files.

Vertical Tab <VT>

Moves the printer to the next vertical tab stop.

The conventional stops for vertical tabs are every eight lines starting at the first printing line on each logical page, that is lines 1, 9, 17, ... within the logical page.

Note that it is difficult to enforce these conventions and it is therefore recommended that vertical tabs not be used in document files.

Back Space <BS>

Moves the printer one character position toward the left edge of the logical page.

Not all these effectors will be used in all format standards, any effectors which are not used in a format standard are ignored.

Page Length

The logical page length will be specified in terms of a number of lines of text.

Page Width

The logical page width will be specified as a number of characters.

Overstriking

overstriking (note that underlining is a subset of overstriking) may be specified to be done in one or both of the following ways, or not at all:

By Line

The composite line is made up of text segments each terminated by the sequence <CR><NUL> except that the final segment is terminated by the sequence <CR><LF>.

By Character

Each character to be overstruck is to be immediately followed by a <BS> and the overstrike character.

End of Line

The end of line convention is the Telnet end of line convention which is the sequence <CR><LF>. It is recommended that use of <CR> and <LF> be avoided in other than the end of line context.

Standard Formats

Format 1 (Basic Document)

This format is designed to be used for documents to be printed on line printers, which normally have 66 lines to a physical page, but often have forced top and bottom margins of 3 lines each.

Active Format Effectors
<FF>, <CR>, <LF>.
Page Length
60 lines.
Page Width
72 Characters.
Overstriking
By Line.

Format 2 (Terminal)

This format is designed to be used with hard copy terminals, which in the normal case have 66 lines to a physical page. It is expected that there are no top or bottom margins enforced by the terminal or its local system, thus any margins around the physical page break must come from the file.

Active Format Effectors
<FF>, <CR>, <LF>, <HT>, <VT>, <BS>.
Page Length
66 lines.
Page Width
72 Characters.
Overstriking
By Character.

Format 3 [Line Printer]

This format is designed to be used with full width (11 by 14 inch paper) line printer output.

Active Format Effectors
<FF>, <CR>, <LF>.
Page Length
60 lines.
Page Width
132 Characters.
Overstriking
None.

Format 4 [Card Image]

This format is designed to be used for simulated card input. The page width is 80 characters, each card image is followed by <CR><LF>, thus each card is represented by between 2 and 82 characters in the file. Note that the trailing spaces of a card image need not be present in the file, and that the early occurence of the <CR><LF> sequence indicates that the remainder of the card image is to contain space characters.

Active Format Effectors
<CR>, <LF>.
Page Length
Infinite.
Page Width
80 Characters.
Overstriking
None.

Format 5 (Center Document)

This format is intended for use with documents to be printed on line printers which normally have 66 lines to the physical page but enforce top and bottom margins of 3 lines each. The text is expected to be centered on the paper. If the horizontal printing density is 10 characters per inch and the paper is 8 and 1/2 inches wide then there will be a one inch margin on each side.

Active Format Effectors
<FF>, <CR>, <LF>.
Page Length
60 Lines.
Page Width
65 Characters.
Overstriking
By Line.

Format 6 (Bound Document)

This format is intended for use with documents to be printed on line printers which normally have 66 lines to the physical page but enforce top and bottom margins of 3 lines each. If the horizontal printing density is 10 characters per inch and the paper is 8 and 1/2 inches wide then the text should be positioned such that there is a 1 and 1/2 inch left margin and a one inch right margin.

Active Format Effectors
<FF>, <CR>, <LF>.
Page Length
60 Lines.
Page Width
60 Characters.
Overstriking
By Line.

Implementation Suggestions

Overflow

Overflow can result from two causes, first if the physical page is smaller than the logical page, and second if the actual text in the file violates the standard under which it is being processed.

In either case the following suggestions are made to implementors of programs which process files in these formats.

Length

If more lines are processed than fit within the minimum of the physical page and the logical page length since the last <FF>, then the <FF> action should be forced.

width

If more character positions are processed than fit on the minimum of the physical page width and the logical page width since the last <CR>, then characters are discarded up to the next <CR>.

or

If more character positions are processed than fit on the minimum of the physical page width and the logical page width since the last <CR>, then the <CR> and <LF> actions should be forced.

References

A. McKenzie "TELNET Protocol Specification," Aug-73, NIC 18639.

"USA Standard Code for Information Interchange," United States of America Standards Institute, 1968, NIC 11246.